Defense Technical Information Center Compilation Part Notice

ADP013506

TITLE: Assessment of Data and Knowledge Fusion Strategies for Diagnostics and Prognostics

DISTRIBUTION: Approved for public release, distribution unlimited

This paper is part of the following report:

TITLE: New Frontiers in Integrated Diagnostics and Prognostics. Proceedings of the 55th Meeting of the Society for Machinery Failure Prevention Technology. Virginia Beach, Virginia, April 2 - 5, 2001

To order the complete compilation report, use: ADA412395

The component part is provided here to allow users access to individually authored sections of proceedings, annals, symposia, etc. However, the component should be considered within the context of the overall compilation report and not as a stand-alone technical report.

The following component part numbers comprise the compilation report:
ADP013477 thru ADP013516

Assessment of Data and Knowledge Fusion Strategies for Diagnostics and Prognostics

Gregory J. Kacprzynski
Michael J. Roemer
Rolf F. Orsagh
Impact Technologies, LLC
125 Tech Park Drive
Rochester, NY 14623
(716) 424-1990

Abstract: Various data, feature and knowledge fusion strategies and architectures have been developed over the last several years for improving upon the accuracy, robustness and overall effectiveness of anomaly, diagnostic and prognostic technologies. Fusion of relevant sensor data, maintenance database information, and outputs from various diagnostic and prognostic technologies has proven effective in reducing false alarm rates, increasing confidence levels in early fault detection, and predicting time to failure or degraded condition requiring maintenance action.

The data fusion strategies discussed in this paper are principally probabilistic in nature and are used to aid in directly identifying confidence bounds associated with specific component fault identifications and predictions. Dempster-Shafer fusion, Bayesian inference, fuzzy-logic inference, neural network fusion and simple weighting/voting are the algorithmic approaches that are discussed in this paper. Data fusion architectures such as centralized fusion, autonomous fusion, and hybrid fusion are described in terms of their applicability to fault diagnosis and prognosis. The final goal is to find the optimal combination of measured system data, data fusion algorithms, and associated architectures for obtaining the highest overall prediction/detection confidence levels associated with a specific application. Evaluation of the fusion and diagnostic strategies was performed using gearbox seeded-fault and accelerated failure data taken with the MDTB (Mechanical Diagnostic Test Bed) at the ARL Lab at Penn State University.

Keywords: Fusion, Feature Extraction, Diagnostics, Prognostics

Introduction: The general objective of data or knowledge fusion is to combine information in the most efficient method possible such that the quality of the fused information is equal to or better than the sum of the parts. Specific to health management, this means reduced uncertainty in current condition assessment reduced (improving diagnostics) and better remaining useful life assessment. Multi-sensor data fusion refers to intelligent processing of an array of 2 or more sensors that have cooperative, complimentary and competitive qualities. As long as the sensor array does not contain independent sensors, arrays usually contain various levels of these three qualities. Cooperative sensors are those that work together to create a new piece of diagnostic information, while a complimentary array creates a more complete picture of a problem. Finally, a competitive array provides unrelated measurements of the same physical phenomena for improved reliability (Brooks, 97).

Fusion Application Areas: Within a health management system, there are three main areas where fusion technologies play a contributing role. These areas are shown in Figure 1. At the lowest level, data fusion can be used to combine information from a multi-sensor data array to validate signals and create features. One example of data fusion is combining a speed signal and a vibration signal to achieve time synchronous averaged vibration features.

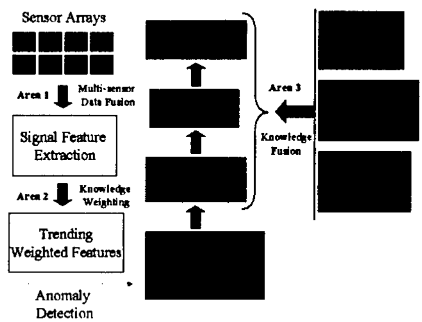

Figure 1 - Fusion Application Areas

At a higher level (area 2), fusion may be used to combine features in intelligent ways so as to obtain the best possible diagnostic information. This would be the case if a feature related to particle count and size in a bearing's lubrication oil were fused with a vibration feature such as kurtosis. The combined result would yield an improved level of confidence about the bearing's health. Finally, Knowledge Fusion (area 3) is used to incorporate experienced-based information such as legacy failure rates or physical model predictions with signal-based information.

One of the main concerns in any fusion technique is the danger of producing a fused system result that is actually performing worse than the best individual tool. This is because poor estimates can drag down the better estimates. The solution to this concern is to weigh the tools according to their capability and performance, which must be realized a priori. The degree of a priori knowledge is a function of the inherent understanding of the physical system and practical experience with the system. The ideal knowledge fusion process for a given application should be selected based on the characteristics of the a priori system information.

Fusion Architectures: Identifying the optimal fusion architecture and approach at each level is a vital factor in assuring that the realized system truly enhances health monitoring capabilities. A brief explanation of fusion architectures will be provided here.

The centralized fusion architecture fuses multi-sensor data while it is still in its raw form as shown in Figure 2. In the fusion center of this architecture, the data is aligned and correlated during the first stage. This means that the competitive or collaborative nature of the data is evaluated and acted upon immediately. Theoretically, this is the most accurate way to fuse data; however, it has the disadvantage of forcing the fusion processor to manipulate a large amount of data. This is often impractical for real-time systems with a relatively large sensor network (Hall, 97).

Centralized Fusion

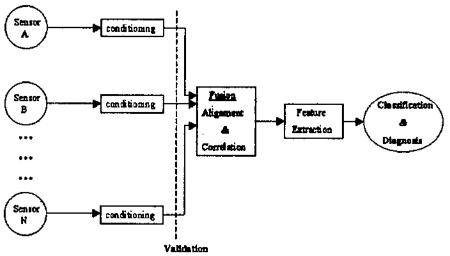

Figure 2 – Centralized Fusion Architecture

The autonomous fusion architecture shown in Figure 3 quells most of the data management problems by placing feature extraction before the fusion process. The creation of features prior to the actual fusion process provides the significant advantage of reducing the dimensionality of the information to be processed. The main undesirable effect of a pure autonomous fusion architecture is that the feature fusion may not be as accurate as in the case of raw data fusion because a significant portion of the raw signal has been eliminated.

Autonomous Fusion

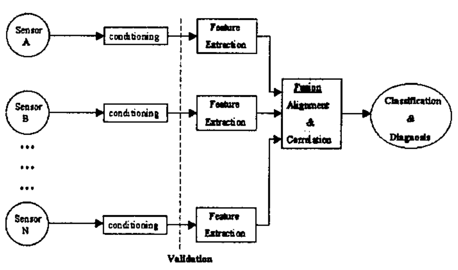

Figure 3 - Autonomous Fusion Architecture

A hybrid fusion architecture takes the best of both and is often considered the most practical because raw data and extracted features can be fused in addition to the ability to "tap" into the raw data if required by the fusion center (Figure 4).

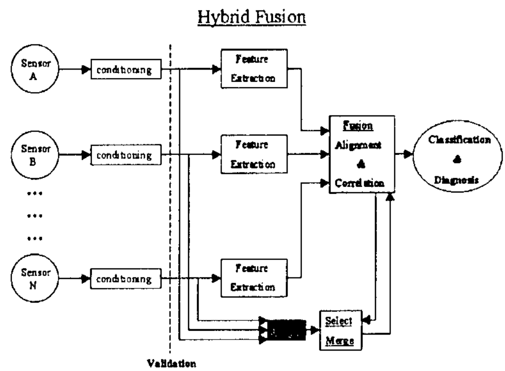

Figure 4 - Hybrid Fusion Architecture

Fusion Techniques: There are probably hundreds of techniques for performing data, feature or knowledge fusion. Because of this fact, sorting through which technique is best can be a daunting and involved task. In addition, there are no hard and fast rules about what fusion techniques or architectures work best for any particular application. The proceeding sections will describe some common fusion approaches such as Baysian Inference, Dempster-Shafer combination, Weighting/Voting, Neural Network Fusion and Fuzzy Logic Inference. A companion paper [3] describes a set of metrics for independently judging the performance and effectiveness of the fusion techniques within a diagnostic system.

Bayesian Inference

Bayesian Inference can be used to determine the probability that a diagnosis is correct, given a piece of a priori information. Analytically this process is described as follows:

$$P(f_1|O_n) = \frac{P(O_n|f_1)\cdot P(f_1)}{\sum_{j=1}^{n} P(O_n|f_1)\cdot P(f_1)} \qquad (1)$$

Where:

$P(f_1|O_n)$ = The probability of fault (f) given a diagnostic output (O), $P(O_n|f_1)$ = The probability that a diagnostic output (O) is associated with a fault (f), and $P(f_1)$ = The probability of the fault (f) occurring.

Bayes' theorem is only able to analyze discrete values of confidence from a diagnostic classifier (i.e. it observes it or it doesn't). Hence, a modified method has been implemented that uses three different sources of information. A-priori probability of failure at time t, ($P_{FO(t)}$) , the probability of failure as determined from the diagnostic classifier ($C_{D(i,t)}$) data, and feature reliability which is independent of time ($R_{D(i)}$). Care must be taken to prevent division by zero.

$$P_{f(t)} = \frac{\sum_{i=1}^{N} C_{D(i,t)}}{P_{FO(t)} \cdot \sum_{i=1}^{N} R_{D(i)}} \qquad (2)$$

The Bayesian process is a common and well established fusion technique, but also has some disadvantages. The knowledge required to generate the a priori probability distributions may not always be available, and instabilities in the process can occur if conflicting data is presented or the number of unknown propositions is large compared to the known propositions.

Dempster-Shafer Method

The Dempster-Shafer method addresses some of the problems discussed above and specifically tackles the a priori probability issue by keeping track of an explicit probabilistic measure of the lack of information. The disadvantage of this method is that the process can become impractical for time critical operations in large fusion problems. Hence, the proper choice of method should be based on the specific diagnostic/prognostic issues that are to be addressed.

In the Dempster-Shafer approach, uncertainty in the conditional probability is considered. The Dempster-Shafer methodology hinges on the construction of a set, called the frame of discernment, which contains every possible hypothesis. Every hypothesis has a belief denoted by a mass probability (m). Beliefs are combined in the following manner.

$$Belief(H_n) = \frac{\sum_{A \cap B = Hn} m_i(A) \cdot m_j(B)}{1 - \sum_{A \cap B = 0} m_i(A) \cdot m_j(B)} \qquad (3)$$

The technique can be best explained through the use of the following example.

Given:

A diagnostic classifier detects Fault A with the following probability and associated uncertainty:
$P_A = 0.80$ +/- 0.15

The a priori probability of Fault A occurring (based on current conditions and a priori information) is the following:
$P_B = 0.30$ +/- 0.10

Therefore: m(A) = 0.65 m(A') = 0.05
m(A,A') = 0.30
m(B) = 0.20 m(B') = 0.60
m(B,B') = 0.20

	B	B'	B,B'
A	0.13	0.39	0.13
A'	0.01	0.03	0.01
A,A'	0.06	0.18	0.06

And: m(A) + m(B) {True} = (0.13 + 0.13 +0.06)/(1-(0.01 +0.39)) = 0.53

This result is called the "belief" and it is the fused probability lower bound. The uncertainty in this result is the following:

$$m(A,A') + m(B,B') = 0.06 / (1-(0.01 + 0.39)) = 0.10 \text{ or } +/- 0.05 \quad (4)$$

Hence, the probability of Fault A having actually occurred given the diagnostic output and in-field experience is 0.58 +/- 0.05.

Weighting/Voting Fusion

Both the Bayesian and Dempster-Shafer techniques can be computationally intensive for real-time applications. A simple weighted average or voting technique is another approach that can be utilized. In both these approaches, weights are assigned based on a prior knowledge of the accuracy of diagnostic/prognostic techniques being used. The only condition is that the sum of the weights must be equal to one. Each confidence value is then multiplied by its respective weight and the results are summed for each moment in time. Weights can also change as a function of time.

$$P(F) = \sum_{n=1}^{i} C_{(i,t)} * W_{(i,t)} \quad (5)$$

Where i is the number of features, C is the confidence value, and W is the weight value for that feature. Although simple in implementation, choosing proper weights is of critical importance to highlighting the proper features under various operating modes.

Fuzzy Logic Inference

Fuzzy Logic Inference is a fusion technique that utilizes the membership function approach to scale and combine specific input quantities to yield a fused output. The basis for the combined output comes from scaling the developed membership functions based on a set of rules developed in a rulebase. Once this scaling is accomplished, the scaled membership functions are combined by one of various methodologies such as summation, maximum or "single best" techniques. Finally, the scaled and combined membership functions are used to calculate the fused output by either taking the centroid, max height or midpoint of the combined function.

An example of a feature fusion process utilizing fuzzy logic is shown below in Figure 5. In this example, features from an image are being combined to help determine if a "foreign" object is present in an original image. Image features such as tonal mean, midtones, kurtosis and many others are combined to give a single output that ranks the probability of an anomalous feature being present in the image.

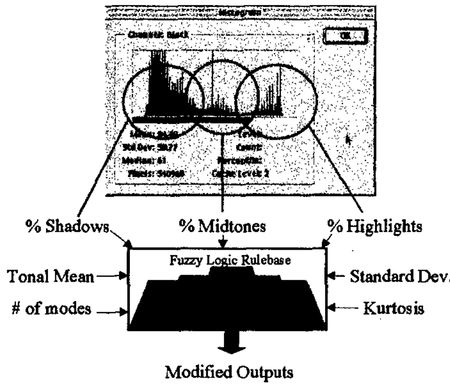

Figure 5 – Example of Fuzzy Logic Inference

Neural Network Fusion

A well accepted application of artificial neural networks (ANNs) is data and feature fusion. For the purposes of fusion, a networks ability to combine information in real-time with the added capability of autonomous re-learning (if necessary) makes it very attractive for many fusion applications.

Artificial neural networks (ANN) utilize a network of simple processing units, each having a small amount of local memory. These units are connected by "communication" links, which carry numerical data. The units operate only on their local data, which is received as input to the units via the connections. Most ANN's have some sort of *training* rule by which the weights of connections are adjusted based on some optimization criterion. Hence, ANN's learn from examples and exhibit certain capability for generalization beyond the training data (examples). ANN's represent a branch of the artificial intelligence techniques that have been increasingly accepted for data fusion and automated diagnostics in a wide range of aerospace applications. Their abilities to fuse features, recognize patterns, and to learn from samples have made ANN's attractive for fusing large data sets from complex systems.

A representative application of neural network fusion would be to combine individual features from different feature extraction algorithms to give a single representative feature. An example of this type of neural network fusion will be given in the following section.

Results

The fusion techniques previously described have been implemented on various vibration features extracted from a data set developed during a series of transitional run-to-failure tests on an industrial gearbox at Penn State ARL. In these tests, the torque was cycled from 100% to 300% load starting at approximately 93 hours. The drivegear experienced multiple broken teeth and the test was stopped at approximately 114 hours. The data collected during the test was processed by many feature extraction algorithm techniques that resulted in 26 vibration features calculated from a single accelerometer attached to the gearbox housing. The features ranged in complexity from a simple RMS level to a measure of the residual signal (gearmesh and sidebands removed) from the time synchronous averaged waveform. More information on these vibration features may be found in [Byington, 1997]. Figures 6 and 7 show plots of two of these features, Kurtosis and NA4, respectively. The smoothed line in each

of these plots is the "ground truth severity" or the probabiltity of failure as determined from visual inspections discussed next.

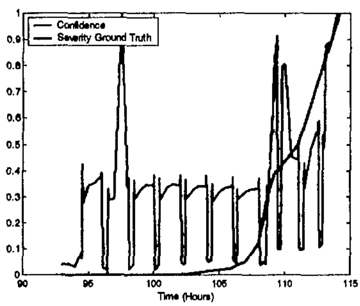

Figure 6 – Kurtosis Feature

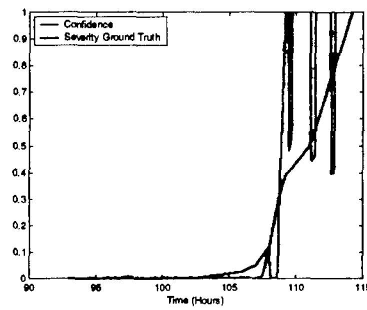

Figure 7 – NA4 Feature

Borescopic inspections of the pinion and drivegear for this particuler test run were performed to bound the time period in which the gear had experienced no damage and when a single tooth had failed. These inspection results, coupled with the evidence of which features were best at detecting tooth cracking prior to failure features (as determined from the diagnsotic metrics discussed later), was the a-priori information used to implement the Bayesian Inference, Weighting/Voting, Neural Network, and Dempster Shafer fusion processes.

The seven best vibration feature as determined by a consistent set of metrics described in [3] were assigned weights of 0.9, average performing features were weighted 0.7, and low performers 0.5 for use in the voting scheme. These weights are directly related to the feature reliability in the Baysian Inference fusion. Similarly, the best features were assigned the uncertainty values of (0.05), average performers (0.10) and low performers (0.15), for the Dempster Shafer combination. The prior probability of failure required for the Neural Network, Bayesian Inference and Dempster Shafer fusion were built upon the experiental evidence that a drive gear crack will form in a mean time of 108 hours with a variance of 2 hours.

Seven of the 26 total vibration features calculated are shown in Figure 8. Note that some of the features have little correlation to the actual tooth failure as defined by the ground truth inspection data. The results of the Dempster-Shafer, Bayesain and Weighted fusion techniques on all 26 features is shown in Figure 9. All three approaches increase in their probability of failure estimates at around 108 hours (index 269). Clearly, the voting fusion is most succeptible to false alarms, the Baysian Inference suggests a probability of failure increase early on but isn't capable of producing a high confidence level. Finally, the Dempster-Shafer combination provides the same early detection, achieves a higher confidence level, but is more sensitive throughout the failure transition region overall.

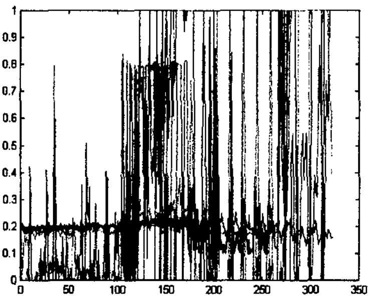

Figure 8 – Top Seven Vibration Features

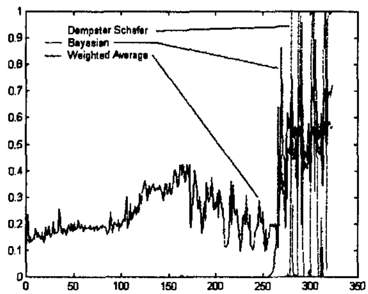

Figure 9 – Fusion of all Features

Next, the same fusion algorithms were applied to just the best seven features. The fusion of these seven features produced more accurate and stable results, which are shown in Figure 10. Note that the Dempster-Shafer combination can now retain a high confidence level with more robustness throughout the critical failure transition region.

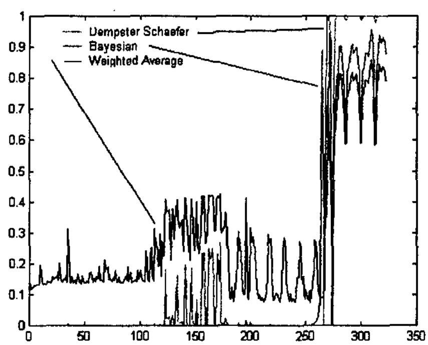

Figure 10 – Fusion of 7 best features

Finally, a simple back propagation neural network was trained on four of the top seven features previously fused (RMS, Kurtosis, NA4, and M8A). In order to train this supervised neural network, the probability of failure as defined by the "ground truth" was required as a-priori information as described earlier. The network automatically adjusts its weights and thresholds (not to be confused with the feature weights) based on the relationships it sees between the probability of failure curve and the correlated feature magnitudes. Figure 11 shows the results of the neural network after being trained by these data sets. The difference between the neural network output and the "ground truth" probability of failure curve is due to error that still exists after the network parameters have optimized to minimize this error. Once trained, the neural network fusion architecture can be used to intelligently fuse these same features for a different test under similar operating conditions.

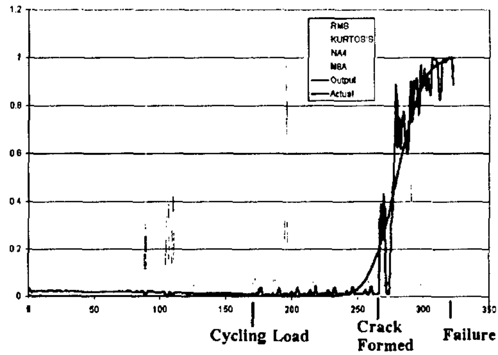

Figure 11 – Fusion with a Neural Network

Conclusion: This paper provides an in-depth discussion about many aspects of fusion including where fusion should exist within a health management system, the different types of fusion architectures, and a number of different fusion techniques. These fusion techniques were applied to vibration features extracted during a transitional failure test associated with an industrial gearbox. The results yielded conclusive evidence that fusion can be very valuable in the diagnostic process if chosen judiciously.

References

[1] Hall, D., and Llinas, J., "An Introduction to Multisensor Data Fusion", Proceedings of the IEEE, January 1997.
[2] Leferve, E., and Colot, O., "A classification method based on the Dempster-Shafer's theory and information criteria" Proceeding of the 2nd International Conference on Information Fusion, July 6-8, 1999.
[3] Orsagh R.F., Roemer M.J., et al "Development of Metrics for Mechanical Diagnostic Technique Qualification and Validation", COMADEM Conference, Houston TX, December 2000.
[4] Agosta, J. M., and Weiss, J. W., "Active Fusion for Diagnosis Guided by Mutual Information Measures", Proceeding of the 2nd International Conference on Information Fusion, July 6-8, 1999
[5] Brooks, R. R., and Iyengar, S. S, Multi-Sensor Fusion, Copyright 1998 by Prentice Hall, Inc., Upper Saddle River, New Jersey 07458
[6] Roemer, M. J. and Kacprzynski, G.J., "Advanced Diagnostics and Prognostics for Gas Turbine Engine Risk Assessment," Paper 2000-GT-30, ASME and IGTI Turbo Expo 2000, Munich, Germany, May 2000.
[7] Roemer, M. J., and Ghiocel, D. M., "A Probabilistic Approach to the Diagnosis of Gas Turbine Engine Faults" Paper 99-GT-363, ASME and IGTI Turbo Expo 1999, Indianapolis, Indiana, June 1999.
[8] Roemer, M. J., and Atkinson, B., "Real-Time Engine Health Monitoring and Diagnostics for Gas Turbine Engines," Paper 97-GT-30, ASME and IGTI Turbo Expo 1997, Orlando, Florida, June 1997.
[9] Byington, C.S., Kozlowski, J.D., "Transitional Data for Estimation of Gearbox Remaining Life", Proceedings of the 51st Meeting of the MFPT, April 1997.